Discrete Mathematics 309 (2009) 5932–5942

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Edge search in graphs with restricted test sets

T. Gerzen

Lehrstuhl II für Mathematik, RWTH Aachen, 52062 Aachen, Germany

ARTICLE INFO

Article history:
Received 5 March 2008
Received in revised form 22 October 2008
Accepted 16 April 2009
Available online 22 May 2009

Keywords:
Edge search
Group testing
Combinatorial search
Information theoretic bound

ABSTRACT

Suppose a graph $G(V, E)$ contains one defective edge e. We search for the endpoints of e by asking questions of the form "Is at least one of the vertices of X an endpoint of e?", where X is a subset of V with cardinality at most p. Then what is the minimum number $c_p(G)$ of questions, which are needed in the worst case to find e?

We solve this search problem suggested by M. Aigner in [M. Aigner, Combinatorial Search, Teubner, 1988] by deriving lower and sharp upper bounds for $c_p(G)$. For the case that G is the complete graph K_n the problem described above is equivalent to the $(2, n)$ group testing problem with test sets of cardinality at most p. We present sharp upper and lower bounds for the worst case number c_p of tests for this group testing problem and show that the maximum difference between the upper and the lower bounds is 3.

1. Introduction

Consider the following classical combinatorial (d, n) sequential group testing problem: assume that a set V of n items contains exactly d defectives. The only way to identify them is through testing. A test can be applied to an arbitrary subset of the n items with two possible outcomes: a negative outcome indicates that all items in the subset are good, a positive outcome indicates that at least one item in the subset is defective. The tests are made one by one, and the outcomes of previous tests are assumed to be known at the time of performing the current test. Let $c_A(d, n)$ denote the maximum (worst case) number of tests required by an algorithm A to identify all defectives. The problem is to determine $c(d, n) = \min_A c_A(d, n)$.

The group testing problem was first proposed by Dorfman (cf. [8]) during World War II when blood samples of millions of draftees were subject to identical analyses in order to detect a few thousand cases of syphilis. In the 1960s Sobel and Groll [14] revived the interest in group testing by giving many industrial applications in detecting chemical leakage and electrical blocking. Over the years many variants of this problem have been discussed in the literature (cf. [4,9,10,1]).

To determine the worst case complexity $c(d, n)$ is an unexpectedly hard problem and is open for $d \geq 2$. The only known general lower bound is the information theoretic bound $\lceil \log_2 \binom{n}{d} \rceil$. Hwang shows in [13] that $c(d, n) \leq \lceil \log_2 \binom{n}{d} \rceil + d - 1$, which is slightly improved in [3] by Allemann.

In this paper we restrict our attention to the case $d = 2$. In this case we can interpret the search domain V as the vertex set of the complete graph K_n. We want to go even further and consider the following generalization of the $(2, n)$ group testing problem: we interpret the search domain V as the vertex set of an arbitrary, finite, simple, undirected graph G with edge set E and search for two defective elements from V, i.e an unknown edge e in E. We write $c(G)$ for the worst case complexity in this case.

A natural restriction on the tests is that only sets of bounded cardinality are allowed. Let us denote by $c_p(G)$ the worst case complexity when only test sets with cardinality at most p are allowed. The extremal case $p = 1$ is investigated by Aigner and Triesch in [2,15] and [16]. The case $p = 2$ was discussed in [12] by Gerzen. In the present paper, we discuss the general case $p \in \mathbf{N}$.

E-mail address: gerzen@math2.rwth-aachen.de.

0012-365X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2009.04.013

For modelling the worst case complexity we introduce the following game-theoretic point of view. We interpret the problem described above as a game between two players A ("Algy") and S ("Strategist") as follows.

Player A asks S questions of the form: "Is at least one of the vertices of $N \subset V$ an endpoint of e?" and receives as answer "yes" or "no". Any sequence of questions determining e is an algorithm of player A. Player S does not fix the edge e at the beginning of the game but delays giving the solution as long as possible. Still the answers she provides to the questions of A have to be consistent, i.e. the graph has to contain an edge which complies with all answers given by S. Any sequence of answers is called a strategy of S. The game stops when e is determined. The p-complexity $c_p(G)$ is the minimum number of questions that has to be asked in order to determine an unknown edge e if both players play optimally.

In the next section we give some notations and known results on c, which we need throughout the paper. In Section 3 we start with the proof of upper bounds for the number of edges and the number of vertices of a graph G with p-complexity c_p. By means of these results we conclude the lower bounds for $c_p(G)$. Section 4 is devoted to the proof of the sharp upper bound for c_p. In Section 5 we consider the complete graph K_n. We provide a sharp lower bound for $c_p(K_n)$ and improve the upper bound of Section 4 for some p. Furthermore, we show that the maximum difference between the upper bound and the lower bound for $c_p(K_n)$ is 3.

2. Preliminaries and terminology

In the following let p, q, m be positive integers with $2 \le p$.

We write $K_{1,m}$ for a star with $m + 1$ vertices. We write $G \cup H = (V(G) \cup V(H), E(G) \cup E(H))$ and $G \setminus H = (V(G) \setminus V(H), E(G) \setminus E(H))$. We obtain the join $G + H$ from $G \cup H$ by adding all edges between G and H. If $W \subset V(G)$, then $G - W = \langle V \setminus W \rangle$ is the subgraph of G obtained by deleting the vertices in W and all edges incident with them. If x and y are non-adjacent vertices of G, then $G + xy$ is obtained from G by joining x to y. We write $N(x)$ for the neighborhood of x in G.

For all graph-theoretic notations which are not defined in the text, we refer the reader to [5] and [6]. The following simple but useful results require no further proof.

Observation 1. *Let G be a graph and let p_1, p_2 be two positive integers with $p_1 \le p_2$. Then*

$$c_{p_1}(G) \ge c_{p_2}(G).$$

Observation 2. *Let G, H be graphs with $H \subset G$. Then*

$$c_p(H) \le c_p(G).$$

Let $c(G)$ be the minimum number of questions in the worst case if there are no restrictions on the test sets. The following proposition is useful throughout the paper.

Proposition 1. *Let $G(V, E)$ be a graph. The information theoretic bound on $c(G)$ is*

$$c(G) \ge \lceil \log_2(|E|) \rceil .$$

A proof can be found for example in [1]. As a simple conclusion we obtain the following corollary.

Corollary 1. *Let $G(V, E)$ be a graph. Then*

$$c_p(G) \ge c(G) \ge \lceil \log_2(|E|) \rceil .$$

For a star with m edges we have the following exact result.

Proposition 2. *Let $q = \lceil \log_2(p) \rceil$. Then*

(a) *$c_p(K_{1,m}) = \lceil \log_2(m) \rceil$ for $m \le 2p$.*
(b) *For $m > 2p$ we get*

$$c_p(K_{1,m}) = \begin{cases} \left\lceil \dfrac{m}{p} \right\rceil + q - 2 & \text{if } 1 \le m \bmod p \le 2^q - p \\ \left\lceil \dfrac{m}{p} \right\rceil + q - 1 & \text{else} \end{cases}$$
$$= t + \lceil \log_2(m - tp) \rceil, \; t = \left\lceil \frac{m}{p} \right\rceil - 2.$$

As conclusions from Observation 2 and Proposition 2 we get the following results.

Lemma 1. *Let $p, r, m_1, \dots, m_r$ be positive integers with $r \le p$ and $m_1 \ge m_2 \ge \dots \ge m_r$. Let $G = K_{1,m_1} \cup K_{1,m_2} \cup \dots \cup K_{1,m_r}$ be an edge disjoint union of stars with pairwise different centers. Then*

$$c_p(K_{1,m_1}) \le c_p(G) \le \lceil \log_2 r \rceil + c_p(K_{1,m_1}) \le 2\lceil \log_2 p \rceil + \left\lceil \frac{m_1}{p} \right\rceil - 1.$$

Lemma 2. *Let $p, r, m_1, \dots, m_r$ be positive integers with $r > p$ and $m_1 \ge m_2 \ge \dots \ge m_r$. Let $G = K_{1,m_1} \cup K_{1,m_2} \cup \dots \cup K_{1,m_r}$ be an edge disjoint union of stars with pairwise different centers. Then*

$$\begin{aligned} c_p(K_{1,m_1}) \le c_p(G) &\le \left\lceil \frac{r}{p} \right\rceil - 1 + \lceil \log_2 p \rceil + c_p(K_{1,m_1}) \\ &\le \left\lceil \frac{r}{p} \right\rceil + 2\lceil \log_2 p \rceil + \left\lceil \frac{m_1}{p} \right\rceil - 2. \end{aligned}$$

For the complexity c Hwang proved the following result in 1972.

Proposition 3 (*Hwang [13]*). *Let $G(V, E)$ be a graph with n vertices. Then*

$$c(G) \le c(K_n) \le \left\lceil \log_2 \binom{n}{2} \right\rceil + 1.$$

In 1994 Damaschke showed that this upper bound can be improved for some graphs.

Proposition 4 (*Damaschke [7]*). *Let $G(V, E)$ be a graph with m edges. Then*

$$c(G) \le \lceil \log_2 m \rceil + 1.$$

For the case $p = 1$ Triesch and Aigner proved in [2] sharp lower bounds for c_1.

Theorem 1 (*Aigner and Triesch [2]*). *Let G be a graph with n vertices and m edges. Then*

(a) $c_1(G) \ge \left\lceil \sqrt{2m - \frac{7}{4}} - \frac{1}{2} \right\rceil$ *and*

(b) $c_1(G) \ge \left\lceil \sqrt{2n - \frac{7}{4}} - \frac{3}{2} \right\rceil$.

For the case $p = 2$ Gerzen proved in [12] sharp upper and lower bounds for c_2.

Theorem 2 (*Gerzen [12]*). *Let G be a graph with $n \ge 6$ vertices and $m \ge 4$ edges. Then*

(a) $c_2(G) \ge \left\lceil \sqrt{\frac{m}{2} - \frac{7}{4}} + \frac{3}{2} \right\rceil$ *and*

(b) $\left\lceil \frac{n+3}{2} \right\rceil \ge c_2(G) \ge \left\lceil \sqrt{\frac{n}{2} - 3} + 1 \right\rceil$.

3. Lower bounds

The information theoretic bound is of course a lower bound for the p-complexity. Let us now look for better lower bounds. First we prove upper bounds for the number of edges and the number of vertices of a graph G with p-complexity c_p. Let $\kappa(G)$ denote the number of components of a graph G.

Theorem 3. *Let $q = \lceil \log_2(p) \rceil$. Let $G(E, V)$ be a graph and $c_p \ge q$ its p-complexity. Then*

(a) $|E| \le p^2 \binom{c_p - q + 2}{2} + 2^q - p^2$ *and*

(b) $|V| \le p^2 \binom{c_p - q + 2}{2} + 2^q - p^2 + \kappa(G) \le p^2 \binom{c_p - q + 2}{2} + 2^q - p^2 + p \cdot c_p$.

Proof. We prove (a) by induction on c_p. For $c_p = q$ the claim follows from Proposition 1. Suppose for $c_p > q$ that A_1 is the first test set in an optimal algorithm for G. It is $|A_1| \le p$ and the inequality

$$d(v) \le p(c_p - q + 1) \tag{1}$$

holds for all vertices $v \in A_1$, where $d(v)$ is the degree of v in G.

Because: Assume that $m = d(v) > p(c_p - q + 1) \geq p$ for a vertex $v \in A_1$ and that player S answers "yes" to the first question. Then the unknown edge e lies in a subgraph of G which contains $K_{1,m}$ as a subgraph. That is why by Observation 2 and Proposition 2 player A needs at least

$$c_p(K_{1,m}) \geq \left\lceil \frac{m}{p} \right\rceil + q - 2 \geq \frac{m}{p} + q - 2 > c_p - 1$$

further questions in the worst case to determine the second end vertex of the unknown edge e. Hence $c_p\,(G) \geq 1 + c_p(K_{1,m}) \geq c_p + 1$, contrary to the assumption.

Now let us consider the subgraph $G - A_1$. The graph $G - A_1$ contains at least one edge because otherwise the choice of A_1 would not be optimal. We get furthermore

$$c_p(G - A_1) \leq c_p - 1. \tag{2}$$

Because: Assume that $c_p(G - A_1) > c_p - 1$ and player S answers "no" to the first question. Then $e \in G - A_1$ and player A needs at least $c_p(G - A_1) > c_p - 1$ further tests in the worst case to determine the second end vertex of the unknown edge e. Thus, we get $c_p(G) \geq c_p(G - A_1) + 1 > c_p$, contrary to the assumption.

We conclude with inequality (1) that

$$|E(G)| \leq |E(G - A_1)| + p \cdot p(c_p - q + 1).$$

Now if $c_p(G - A_1) \geq q$ we use the induction and the inequality (2) for the first term of the sum to get

$$\begin{aligned} |E(G)| &\leq p^2 \binom{c_p - 1 - q + 2}{2} + 2^q - p^2 + p^2(c_p - q + 1) \\ &= p^2 \binom{c_p - q + 2}{2} + 2^q - p^2. \end{aligned}$$

If $c_p(G - A_1) < q$, then $|E(G - A_1)| < 2^q \leq p^2 \binom{c_p - 1 - q + 2}{2} + 2^q - p^2$ by Corollary 1 and we receive again

$$|E(G)| \leq p^2 \binom{c_p - q + 2}{2} + 2^q - p^2.$$

The first inequality in (b) follows immediately from (a). Observe for the second inequality that the number of components of G is bounded by $p \cdot c_p$ since G has no isolated vertices. ■

Solving the inequalities in Theorem 3 we obtain the lower bounds for c_p.

Corollary 2. *Let $q = \lceil \log_2(p) \rceil$. Let G be a graph with n vertices and m edges. Then*

(a) *either $c_p(G) < q$ or $c_p(G) \geq \frac{2q-3}{2} + \sqrt{\frac{9}{4} - \frac{4p-2m}{p^2}}$ and*

(b) *either $c_p(G) < q$ or $c_p(G) \geq q - \frac{3p+2}{2p} + \sqrt{\frac{9}{4} + \frac{3-2q}{p} + \frac{1+2n-4p}{p^2}}$.*

Proof of Corollary 2. Solving the inequality (a) in Theorem 3 we get

$$c_p(G) \geq \frac{2q-3}{2} + \sqrt{\frac{9}{4} - \frac{2^{q+1} - 2m}{p^2}}$$

and therefore

$$c_p(G) \geq \frac{2q-3}{2} + \sqrt{\frac{9}{4} - \frac{4p - 2m}{p^2}}$$

since $2^{q+1} \leq 4p$.

Solving the inequality (b) in Theorem 3 we get

$$c_p(G) \geq q - \frac{3p+2}{2p} + \sqrt{\frac{9}{4} + \frac{3-2q}{p} + \frac{1 + 2n - 2^{q+1}}{p^2}}$$

and thus

$$c_p(G) \geq q - \frac{3p+2}{2p} + \sqrt{\frac{9}{4} + \frac{3-2q}{p} + \frac{1 + 2n - 4p}{p^2}}$$

again since $2^{q+1} \leq 4p$. ■

Therefore, we obtain by means of Corollary 1 the following corollary.

Corollary 3. *Let $q = \lceil \log_2(p) \rceil$. Let G be a graph with $n > 2^{q+1} + 1$ vertices and $m > 2^q$ edges. Then*

(a) $c_p(G) \geq q - \frac{3}{2} + \sqrt{\frac{9}{4} - \frac{4p-2m}{p^2}}$ *and*

(b) $c_p(G) \geq q - \frac{3p+2}{2p} + \sqrt{\frac{9}{4} + \frac{3-2q}{p} + \frac{1+2n-4p}{p^2}}$.

Observation 3. *Let G be a graph with m edges. For $p \geq 4$, $q = \lceil \log_2 p \rceil$ and $m \geq 2^q$ we estimate $\frac{2q-3}{2} + \sqrt{\frac{9}{4} - \frac{4p-2m}{p^2}} \geq \sqrt{\frac{9}{4} - \frac{4p-2m}{p^2}}$ and hence*

$$\frac{2q-3}{2} + \sqrt{\frac{9}{4} - \frac{2^{q+1} - 2m}{p^2}} \geq \lceil \log m \rceil \quad \textit{if} \ \frac{2m}{p^2} - \lceil \log m \rceil^2 \geq \frac{4}{p} - \frac{9}{4}.$$

Therefore, we conclude that the lower bound of Corollary 3 is better than the information theoretic bound if

$$\frac{2m}{p^2} - \lceil \log_2 m \rceil^2 \geq \frac{4}{p} - \frac{9}{4}.$$

We have seen sharp lower bounds for $c_1(G)$ and $c_2(G)$ in Section 2. The upper bound for $|E(G)|$ of Theorem 3 is not sharp.

4. Upper bounds

Let us now turn our attention to upper bounds for c_p. For the case that there are no restrictions on the test sets we have the upper bound $\lceil \log_2 n(n-1) \rceil$ from Hwang. This implies the following result for the complete graph K_n.

Lemma 3. *Let $n \leq 2p + 1$. Then*

$$c_p(K_n) \leq \lceil \log_2 n(n-1) \rceil .$$

Proof. From Proposition 3 it follows that $c(K_n) \leq \lceil \log_2 n(n-1) \rceil$, where $c(K_n)$ is the complexity of K_n with no restriction on the test sets. In the case $n \leq 2p + 1$ the restriction $|X| \leq p$ on a test set X is irrelevant. ■

In the next sections complete split graphs will play an important role. First we need the definition of a split graph (see also [11]).

Definition 1. *A graph G is a split graph if there is a partition of its vertex set into two nonempty subsets V and W such that V induces a complete graph and W induces an empty graph. We say that V is the clique set of G and W is the independent set of G.*

Definition 2. *A split graph G with clique set V and independent set W is called complete if every vertex of V is adjacent to every vertex of W. A complete split graph with $|V| = p$ and $|W| = n - p$ is called $n - p$-graph with notation $G = H_{n,p}$.*

Proposition 5. *Let $n \geq 2p$. Then*

$$c_p(K_n) \leq 2 \lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil .$$

Proof. For $n = 2p$ and $n = 2p + 1$ the claim follows from Lemma 3.

For $n = 2p + 1$ player A probes a set of p arbitrary vertices of K_n. If player S answers "yes" player A has to search in $H_{n,p}$. Note that the $n - p$-graph $H_{n,p}$ is a subgraph of the graph $\cup_{i=1}^{p} K_{1,n-1} = \cup_{i=1}^{p} K_{1,2p}$. If S answers "no" the unknown edge e lies in the complete graph K_{p+1}. Thus, we get by Observation 2

$$\begin{aligned} c_p(K_n) &\leq \max\left\{1 + c_p(H_{n,p}),\ 1 + c_p(K_{p+1})\right\} \\ &\leq \max\left\{1 + c_p(\cup_{i=1}^{p} K_{1,n-1}),\ 1 + c_p(K_{p+1})\right\} . \end{aligned}$$

Now using Lemma 1 for the first term and Lemma 3 for the second term we obtain

$$c_p(K_n) \leq 1 + \max\left\{\lceil \log_2 p \rceil + c_p(K_{1,2p}),\ \lceil 2 \log_2(p+1) \rceil\right\}$$

and thus by Proposition 2

$$\begin{aligned} c_p(K_n) &\leq 1 + \max\{\lceil \log_2 p \rceil + \lceil \log_2 2p \rceil,\ \lceil 2 \log_2(p+1) \rceil\} \\ &\leq 2 \lceil \log_2 p \rceil + 2 = 2 \lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil . \end{aligned}$$

For $n \geq 2p+2$ we proceed by induction on n. Player A first probes a set consisting of arbitrary p vertices of K_n. If player S answers "yes", player A has to search in $H_{n,p}$. Note that the $n-p$-graph $H_{n,p}$ is a subgraph of the graph $\cup_{i=1}^{p} K_{1,n-1}$. If S answers "no" the unknown edge e lies in the complete graph K_{n-p}. Thus, we deduce by Observation 2 that

$$c_p(K_n) \leq \max\left\{1 + c_p(\cup_{i=1}^{p} K_{1,n-1}),\, 1 + c_p(K_{n-p})\right\}.$$

Now using Lemma 1 for the first term and induction for the second term if $n-p \geq 2p$ we obtain

$$c_p(K_n) \leq 1 + \max\left\{\lceil \log_2 p\rceil + c_p(K_{1,n-1}),\, 2\lceil \log_2 p\rceil + \left\lceil \frac{n-p-1}{p}\right\rceil\right\}$$

and thus by Proposition 2

$$\begin{aligned} c_p(K_n) &\leq 1 + \max\left\{\lceil \log_2 p\rceil + \lceil \log_2(p)\rceil + \left\lceil \frac{n-1}{p}\right\rceil - 1,\, 2\lceil \log_2 p\rceil + \left\lceil \frac{n-p-1}{p}\right\rceil\right\} \\ &= 2\lceil \log_2 p\rceil + \left\lceil \frac{n-1}{p}\right\rceil. \end{aligned}$$

If $p+2 \leq n-p < 2p$, then by Lemma 3 we estimate $c_p(K_{n-p}) \leq c_p(K_{2p}) \leq \lceil 2\log_2 2p\rceil$ and receive again by means of Proposition 2

$$\begin{aligned} c_p(K_n) &\leq 1 + \max\left\{\lceil \log_2 p\rceil + c_p(K_{1,n-1}),\, \lceil 2\log_2 2p\rceil\right\} \\ &\leq 1 + \max\left\{\lceil \log_2 p\rceil + \lceil \log_2(p)\rceil + \left\lceil \frac{n-1}{p}\right\rceil - 1,\, \lceil 2\log_2 p\rceil + 2\right\} \\ &\leq 2\lceil \log_2 p\rceil + \left\lceil \frac{n-1}{p}\right\rceil. \quad \blacksquare \end{aligned}$$

From Lemma 3 and Proposition 5 we deduce

Theorem 4. *Let G be a graph with n vertices. Then*

$$c_p(G) \leq 2\lceil \log_2 p\rceil + \left\lceil \frac{n-1}{p}\right\rceil.$$

To see that the upper bound of Theorem 4 is sharp let us focus on the complete graph.

5. The complete graph K_n

To find an unknown edge in K_n is the same as identifying 2 defective elements in a set of n items through testing. Thus, the p-complexity $c_p(K_n)$ is exactly the minimum number of tests for the (2,n) group testing problem with test sets of cardinality at most p.

In this section we will work out a sharp lower bound for $c_p(K_n)$ and we will see that the upper bound of Proposition 5 is sharp for an arbitrary p and $n \geq 4p$, but can be improved for some p.

Suppose that player A probes a set of $k \leq p$ vertices of K_n. The resulting graphs are the complete graph K_{n-k} and the n–k-graph $H_{n,k}$.

Therefore, it is a better strategy for player S to answer "no" as long as the inequality $|E(K_{n-p})| > |E(H_{n,p})|$ holds. It is

$$|E(K_{n-p})| = \binom{n-p}{2} = \frac{n^2 - 2np - n + p^2 + p}{2}$$

and

$$|E(H_{n,p})| = \binom{p}{2} + p(n-p) = \frac{2np - p - p^2}{2}.$$

Thus, the inequality $|E(K_{n-p})| > |E(H_{n,p})|$ holds for all $n \geq 4p$. Now we can use this to deduce a good lower bound for $c_p(K_n)$.

Proposition 6. *Let $n \geq 4p$. Then*

$$c_p(K_n) \geq \lceil \log_2(3p(3p-1))\rceil + \left\lfloor \frac{n}{p}\right\rfloor - 4.$$

Proof. Consider the following strategy of player S: Suppose that after i tests player A knows that the unknown edge e lies in the graph $G_i\,(V_i, E_i)$ where $G_0 = K_n$. As long as the inequality $|V_i| \geq 4p$ holds, player S answers "no" in the $(i+1)$th test. Then G_{i+1} is a complete graph with at least $n-(i+1)p$ vertices as long as $n - ip \leq 4p$. For $j = \left\lfloor \frac{n-4p}{p} \right\rfloor$ we get

$$|V_j| \geq |V_0| - jp \geq 4p \text{ and } |V_{j+1}| \geq 3p.$$

Thus, we obtain $G_{j+1} = K_k$ with $k = |V_{j+1}| \geq 3p$ and $c_p(K_n) \geq j + 1 + c_p(K_k)$. Together with Proposition 1 this yields

$$\begin{aligned} c_p(K_n) &\geq \left\lfloor \frac{n-4p}{p} \right\rfloor + 1 + c_p(K_k) \\ &\geq \left\lfloor \frac{n}{p} \right\rfloor - 3 + \left\lceil \log_2 \binom{k}{2} \right\rceil \\ &\geq \lceil \log_2(3p(3p-1)) \rceil + \left\lfloor \frac{n}{p} \right\rfloor - 4. \quad \blacksquare \end{aligned}$$

Now we want to see that the upper bound of Theorem 4 is sharp. Hence let us look for the difference $d_{n,p}$ between this upper bound and the lower bound of Proposition 6. Let $q = \lceil \log_2(p) \rceil$. For $n \geq 4p$ we get

$$\begin{aligned} d_{n,p} &:= 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - \lceil \log_2(3p(3p-1)) \rceil - \left\lfloor \frac{n}{p} \right\rfloor + 4 \\ &\leq 2q + \left\lceil \frac{n-1}{p} \right\rceil - \left\lceil \frac{n-1}{p} \right\rceil + 5 - \lceil \log_2(3p(3p-1)) \rceil \\ &\leq 2q + 5 - \left\lceil \log_2 3 + \log_2\left(2^{q-1}\right) + \log_2\left(3(2^{q-1}+1) - 1\right) \right\rceil \\ &\leq 2q + 5 - \left\lceil \log_2 3 + q - 1 + \log_2 3 + \log_2 2^{q-1} \right\rceil \\ &\leq 2q + 5 - \left\lceil \log_2 3 + q - 1 + \log_2 3 + \log_2 2^{q} - 1 \right\rceil \\ &\leq 7 - \lceil 2\log_2 3 \rceil = 3. \end{aligned}$$

Observation 4. *Thus, we obtain that the inequalities*

$$\lceil \log_2(3p(3p-1)) \rceil + \left\lfloor \frac{n}{p} \right\rfloor - 4 \leq c_p(K_n) \leq 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil$$

hold for $n \geq 4p$ and the maximum difference between the upper and the lower bound for $c_p(K_n)$ is 3.

In the case $p = 2^q \geq 4$ we even get that the maximum difference is 0 or 1 depends on n:

$$\begin{aligned} d_{n,p} &= 2q + \left\lceil \frac{n-1}{p} \right\rceil - \left\lceil \log_2 3 + \log_2 2^q + \log_2(3 \cdot 2^q - 1) \right\rceil - \left\lfloor \frac{n}{p} \right\rfloor + 4 \\ &\leq q + \left\lceil \frac{n-1}{p} \right\rceil - \left\lfloor \frac{n}{p} \right\rfloor + 4 - \left\lceil \log_2 3 + \log_2\left(3 \cdot 2^q - 1\right) \right\rceil \\ &= \left\lceil \frac{n-1}{p} \right\rceil - \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor q + 4 - \log_2 3 - \log_2\left(3 \cdot 2^q - 1\right) \right\rfloor \\ &= \left\lceil \frac{n-1}{p} \right\rceil - \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor 4 - \log_2 3 - \log_2\left(3 - \frac{1}{2^q}\right) \right\rfloor \\ &\leq \left\lceil \frac{n-1}{p} \right\rceil - \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor 4 - \log_2 3 - \log_2\left(3 - \frac{1}{4}\right) \right\rfloor \\ &= \left\lceil \frac{n-1}{p} \right\rceil - \left\lfloor \frac{n}{p} \right\rfloor + \lfloor 6 - \log_2 33 \rfloor \\ &= \begin{cases} 0, & \text{if } n \equiv 1 \text{or } n \equiv 0 \bmod p \\ 1 & \text{else.} \end{cases} \end{aligned}$$

Therefore, we can deduce the following corollary.

Corollary 4. *Let $p = 2^q \geq 4$ and $n \geq 4p$ be positive integers with $n \equiv 1$ or $n \equiv 0 \bmod p$. Then*

$$c_p(K_n) = 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil.$$

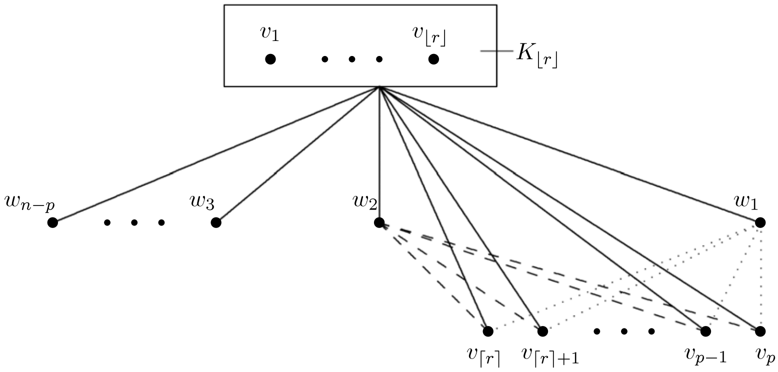

Fig. 1. The graph H_1.

Observation 5. *We have seen that both the upper bound of Theorem 4 as well as the lower bound of Proposition 6 are sharp. But we have no characterization for all graphs with equality in Theorem 4 or in Proposition 6.*

We just considered the case that p is a power of 2. Let us now look at the other extreme case $p = 2^q + 1$. We are going to see that in this case it is possible to improve the upper bound of Proposition 5.

Proposition 7. *Let $p = 2^q + 1$ and $n \geq 2p + 2$. Then*

$$c_p(K_n) \leq 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 1.$$

We need a preliminary Lemma.

Lemma 4. *Let $p = 2^q + 1$ and $n \geq 2p + 2$. Let $H_{n,p}$ be the $n - p$-graph. Then*

$$c_p(H_{n,p}) \leq 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 2.$$

Proof of the Lemma. Let $r = \frac{p}{2}$. Let $V = \{v_1, \dots, v_p\}$ be the clique set of $H_{n,p}$ and $W = \{w_1, \dots, w_{n-p}\}$ the independent set of $H_{n,p}$.

In the first test player A probes the set $A_1 = \{v_1, \dots, v_{\lfloor r \rfloor}, w_1, w_2\}$. If S answers "no" in the first test, player A has to search in the graph $G_1 = H_{n-2-\lfloor r \rfloor, \lceil r \rceil}$ with clique set $\{v_{\lceil r \rceil}, \dots, v_p\}$ and independent set $\{w_3, \dots, w_{n-p}\}$. If the answer of S is "yes", the unknown edge e belongs to the graph $H_1 = H_{n,p} - E(G_1)$.

Let us consider the case that S answers "yes" in the first test. The resulting graph H_1 has the vertex set $V \cup W$ and edge set

$$E(H_1) = \{v_i v_j, v_i w_l | i = 1, \dots, \lfloor r \rfloor; j = 1, \dots, p; l = 1, \dots, n-p\} \cup \{w_1 v_j, w_2 v_j | j = \lceil r \rceil, \dots, p\}.$$

Thus H_1 comprises of the graph $H_{n,\lfloor r \rfloor}$ (with clique set $\{v_1, \dots, v_{\lfloor r \rfloor}\}$ and independent set $\{w_1, \dots, w_{n-p}, v_{\lceil r \rceil}, \dots, v_p\}$) and two stars $K_{1,\lfloor r \rfloor}$ joining the centers w_1 or w_2 with the vertices $v_{\lceil r \rceil}, \dots, v_p$ (see Fig. 1).

Player A now probes the set $A_2 = \{w_1, \dots, w_p\}$. If player S answers "yes", then the unknown edge e lies in the graph H_2 with vertex set $V \cup \{w_1, \dots, w_p\}$ and edge set

$$E(H_2) = \{v_i w_l | i = 1, \dots, \lfloor r \rfloor; l = 1, \dots, p\} \cup \{w_1 v_j, w_2 v_j | j = \lceil r \rceil, \dots, p\}.$$

Hence, H_2 is the edge disjoint union $\cup_{j=1}^{\lfloor r \rfloor} K_{1,p} \cup K_{1,\lfloor r \rfloor} \cup K_{1,\lfloor r \rfloor}$ where the center vertices of the first $\lfloor r \rfloor$ stars are $v_1, \dots, v_{\lfloor r \rfloor}$ and the centers of the last two stars are the vertices w_1, w_2 (see Fig. 2). Therefore, A needs at most $\lceil \log_2(\lfloor r \rfloor + 2) \rceil + \lceil \log_2 p \rceil$ further tests to find e by Proposition 2 and Lemma 1. If player S answers "no", then e lies in $\tilde{H_2} = H_{n-p,\lfloor r \rfloor}$ with clique set $\{v_1, \dots, v_{\lfloor r \rfloor}\}$ and star set $\{w_{p+1}, \dots, w_{n-p}\}$. We note that $\tilde{H_2} \subset \cup_{j=1}^{\lfloor r \rfloor} K_{1,n-1-p}$ and thus, by Proposition 2 and Lemma 1

$$\begin{aligned} c_p(\tilde{H_2}) &\leq \left\lceil \log_2 \left\lfloor \frac{p}{2} \right\rfloor \right\rceil + \left\lceil \frac{n-p-1}{p} \right\rceil + \lceil \log_2 p \rceil - 1 \\ &= \left\lceil \log_2 \left\lfloor \frac{2^q+1}{2} \right\rfloor \right\rceil + \left\lceil \frac{n-1}{p} \right\rceil + \lceil \log_2 p \rceil - 2 \\ &= 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 4. \end{aligned}$$

Note that the equality holds only for $p = 2^q + 1$.

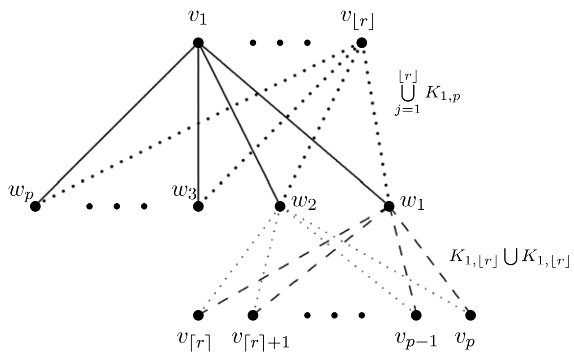

Fig. 2. The graph H_2.

In total we get

$$\begin{aligned}
c_p(H_1) &\leq \max\left\{1 + c_p(H_2),\, 1 + c_p(\tilde{H}_2)\right\} \\
&\leq \max\left\{1 + \left\lceil \log_2\left(\left\lfloor \frac{2^q+1}{2} \right\rfloor + 2\right)\right\rceil + \lceil \log_2 p \rceil,\, 1 + 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 4\right\} \\
&= \max\left\{1 + \lceil \log_2(2^{q-1}+2) \rceil + \lceil \log_2 p \rceil,\, 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 3\right\} \\
&= \max\left\{2\lceil \log_2 p \rceil,\, 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 3\right\} \\
&= 2\lceil \log_2(p) \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 3.
\end{aligned}$$

The last equality holds, since $n \geq 2p+2$.

Let us consider now the case that player S answers "no" in the first test. A has to search in the graph $G_1 = H_{n-2-\lfloor r \rfloor, \lceil r \rceil} = H_{n-2-2^{q-1}, 2^{q-1}+1}$ with clique set $V_1 = \{v_{\lceil r \rceil}, \ldots, v_p\} = \{v_{2^{q-1}+1}, \ldots, v_p\}$. Player A probes a set of $\left\lfloor \frac{|V_1|}{2} \right\rfloor = \left\lfloor \frac{2^q+1-2^{q-1}}{2} \right\rfloor = 2^{q-2}$ arbitrary vertices of the set V_1 in the first test. Suppose for $i \geq 1$ that after i tests player A knows for sure that the unknown edge e lies in the subgraph G_{i+1} of G_1. As long as player S answers "no" in the ith test, the resulting graph after i tests is

$$G_{i+1} = H_{n-2-(2^{q-1}+\cdots+2^{q-i-1}), 2^{q-i-1}+1}$$

and player A probes a set of $\left\lfloor \frac{|V_{i+1}|}{2} \right\rfloor = 2^{q-i-1}$ arbitrary vertices of the clique set V_{i+1} of G_{i+1} in the next test.

If S answers always "no", then after $\lceil \log_2 p \rceil - 2 = q - 1$ tests player A knows that the unknown edge e lies in the graph

$$G_q = H_{n-2-(2^{q-1}+\cdots+2^{q-(q-1)-1}), 2^{q-(q-1)-1}+1} = H_{n-2-(2^q-1), 2} \subset K_{1,n-2-2^q} \cup K_{1,n-2-2^q}.$$

Therefore, A needs at most

$$1 + c_p(K_{1,n-2-2^q}) \leq 1 + \lceil \log_2 p \rceil + \left\lceil \frac{n-2-2^q}{p} \right\rceil - 1 = \lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 1$$

tests to find e in G_q by Proposition 2 and Lemma 1. We conclude that A needs at most

$$q - 1 + \lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 1 = 2\lceil \log_2 p \rceil - 3 + \left\lceil \frac{n-1}{p} \right\rceil$$

tests to identify e in the graph G_1 if S answers always "no".

If in one test player S answers "yes", then we get

$$G_k = H_{n-2-(2^{q-1}+\cdots+2^{q-k}), 2^{q-k}+1} \quad \text{and} \quad G_{k+1} = H_{n-2-(2^{q-1}+\cdots+2^{q-k}), 2^{q-k-1}},$$

where $k \leq q$ is the first test in which S answers "yes". Let $a := n - 3 - (2^{q-1} + \cdots + 2^{q-k}) = n - 3 - 2^q + 2^{q-k}$. Note that $a \geq p + 2$ for $n \geq 2p + 2$. A knows that e lies in $G_{k+1} \subset \bigcup_{j=1}^{2^{q-k-1}} K_{1,a}$ and thus needs at most $\lceil \log_2 2^{q-k-1} \rceil + c_p(K_{1,a})$ tests to find e in G_{k+1} by Lemma 1. Let us consider the star $K_{1,a}$. By Proposition 2 we know that

$$\begin{aligned} c_p(K_{1,a}) &\le \left\lceil \frac{a}{p} \right\rceil + \lceil \log_2 p \rceil - 1 \\ &= \left\lceil \frac{n-3-2^q+2^{q-k}}{p} \right\rceil + \lceil \log_2 p \rceil - 1 \\ &\le \left\lceil \frac{n-1}{p} \right\rceil + \lceil \log_2 p \rceil - 1. \end{aligned}$$

Therefore, we can estimate

$$\begin{aligned} c_p(G_{k+1}) &\le \left\lceil \log_2 2^{q-k-1} \right\rceil + c_p(K_{1,a}) \\ &\le q-1-k+\left\lceil \frac{n-1}{p} \right\rceil + \lceil \log_2 p \rceil - 1 \\ &= 2\lceil \log_2(p) \rceil - 3 - k + \left\lceil \frac{n-1}{p} \right\rceil. \end{aligned}$$

Hence, we obtain

$$\begin{aligned} c_p(G_1) &\le \max\left\{ 2\lceil \log_2 p \rceil - 3 + \left\lceil \frac{n-1}{p} \right\rceil, k + c_p(G_{k+1}) \right\} \\ &\le \max\left\{ 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 3, k + 2\lceil \log_2 p \rceil - 3 - k + \left\lceil \frac{n-1}{p} \right\rceil \right\} \\ &= 2\lceil \log_2(p) \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 3. \end{aligned}$$

In total we receive

$$\begin{aligned} c_p(H_{n,p}) &\le \max\left\{ 1 + c_p(H_1), 1 + c_p(G_1) \right\} \\ &= 2\lceil \log_2(p) \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 2. \end{aligned}$$

This completes the proof of the lemma. ■

Proof of Proposition 7. Let us consider the case $n \le 3p+1$ first. Player A probes the set consisting of arbitrary p vertices of K_n first. Then the resulting graphs are $G_1 = H_{n,p}$ and $G_2 = K_{n-p}$ with $n - p \le 2p+1$. Thus, we obtain by Lemmas 3 and 4 that

$$\begin{aligned} c_p(K_n) &\le \max\left\{ 1 + c_p(H_{n,p}), 1 + c_p(K_{n-p}) \right\} \\ &\le \max\left\{ 1 + 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 2, 1 + \lceil 2\log_2(n-p) \rceil \right\} \\ &= 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 1. \end{aligned}$$

The last equality holds for $2p+2 \le n \le 3p+1$ since:

$$\begin{aligned} & 1 + \lceil 2\log_2(n-p) \rceil - \left(1 + 2\lceil \log_2 p \rceil + \left\lceil \frac{n-1}{p} \right\rceil - 2 \right) \\ &= \lceil 2\log_2(n-p) \rceil - 2q - \left\lceil \frac{n-1}{p} \right\rceil = \lceil 2\log_2(n-p) \rceil - 2q - 3 \\ &\le \lceil 2\log_2(2p+1) \rceil - 2q - 3 = \left\lceil 2\log_2 \frac{2^{q+1}+3}{2^q} \right\rceil - 3 \\ &\le \left\lceil 2\log_2\left(2 + \frac{3}{4} \right) \right\rceil - 3 = 0 \text{ for } q \ge 2. \end{aligned}$$

For $p = 3$ (i.e. $q = 1$) it is easy to see that $c_3(K_8) = 6 = 2\lceil \log_2 3 \rceil + \lceil 7/3 \rceil - 1$, $c_3(K_9) = 6 = 2\lceil \log_2 3 \rceil + \lceil 8/3 \rceil - 1$ and $c_3(K_{10}) = 6 = 2\lceil \log_2 3 \rceil + \lceil 9/3 \rceil - 1$.

For $n \geq 3p + 2$ we proceed by induction on n. Player A probes the set consisting of arbitrary p vertices of K_n first. Then the resulting graphs are $G_1 = H_{n,p}$ and $G_2 = K_{n-p}$ with $n - p \geq 2p + 2$. Thus, we deduce by induction and Lemma 4 that

$$\begin{aligned} c_p(K_n) &\leq \max\left\{1 + c_p(H_{n,p}),\, 1 + c_p(K_{n-p})\right\} \\ &\leq \max\left\{1 + 2\lceil \log_2 p\rceil + \left\lceil \frac{n-1}{p} \right\rceil - 2,\, 1 + 2\lceil \log_2 p\rceil + \left\lceil \frac{n-p-1}{p} \right\rceil - 1\right\} \\ &= 2\lceil \log_2 p\rceil + \left\lceil \frac{n-1}{p} \right\rceil - 1. \end{aligned}$$

This completes the proof. ■

References

[1] M. Aigner, Combinatorial Search, Teubner, 1988.
[2] M. Aigner, E. Triesch, Searching for an edge in a graph, J. Graph Theory 12 (1988) 45–57.
[3] A. Allemann, Improved upper bounds for several variants of group testing, Ph.D. thesis, RWTH Aachen, Germany, 2003.
[4] R. Alswede, I. Wegener, Suchprobleme, Teubner, Stuttgart, 1979.
[5] B. Bollobas, Modern Graph Theory, Springer, 1998.
[6] G. Chartrand, P. Zhang, Introduction to Graph Theory, McGraw-Hill, 2005.
[7] P. Damaschke, A tight upper bound for group testing in graphs, Discrete. Appl. Math. 48 (1994) 101–109.
[8] R. Dorfman, The detection of defective members of large population, Ann. Math. Statistics 14 (1943) 436–440.
[9] D.Z. Du, F.K. Hwang, Combinatorial group testing and its applications, Word Scientific, 1993.
[10] D.Z. Du, F.K. Hwang, Pooling Designs And Nonadaptive Group Testing: Important Tools For DNA Sequencing, Word Scientific, 2006.
[11] S. Fördes, P.L. Hammer, Split graphs, Congr. Numer. 19 (1977) 311–315.
[12] T. Gerzen, Searching for an edge in a graph with restricted test sets, Discrete Math. 309 (6) (2009) 1334–1346.
[13] F.K. Hwang, A method for detecting all defective members in a population by group testing, J. Amer. Statist. Assoc. 67 (1972) 605–608.
[14] M. Sobel, P.A. Groll, Group testing to estimate efficiently all detectives in a binomial sample, Bell Syst. Tech. J. 28 (1959) 1179–1252.
[15] E. Triesch, On a search problem in graph theory, optimization-fifth French–German conference castel novel 1988, in: Lecture Notes in Mathematics 1405, Springer, 1989, pp. 171–176.
[16] E. Triesch, A probabilistic upper bound for the edge identification complexity of graphs, Discrete Math. 125 (1994) 371–376.